ISO 19115-3 standard implementation in geometa R package

Emmanuel Blondel

2024-04-01

Signatories

Project team

Emmanuel Blondel

- **ORCID**: https://orcid.org/0000-0002-5870-5762)
- **Website**: https://eblondel.github.io
- **Github**: https://github.com/eblondel
- **r-universe**: https://eblondel.r-universe.dev
- **Presentation**: Author of several R CRAN packages focusing on geospatial such as geometa, ows4R, geosapi, geonapi, and related to (meta)data management in general such as atom4R, zen4R, ocs4R, rsdmx. Contributor/developer to Java/Python/JavaScript spatial libraries from the OSGeo community, related to ISO/OGC standards. Graduated engineer in Agronomic and Environmental Sciences (2008, ENSAT, Toulouse, France), specialized in GIS with focus on promoting and implementing international standards and fostering open science. Currently consultant (individual micro-company), senior GIS and R expert for the Food and Agriculture Organization of the United Nations (UN-FAO) and World Bank, international expert at UN for marine geospatial information management. Collaborates actively with scientific institutes and projects focusing on methodologies based on open geospatial (meta)data standards, and data access & exploration tools enabling open and reproducible science.

Contributors

Contributors that have participated to project review:

- Mr Anton Ellenbroek - Information systems analyst at UN-FAO - Contact: Anton.Ellenbroek@fao.org

Consulted

OSGeo development community

- Mr Tom Kralidis - Senior Geospatial Architect at Meteorological Service of Canada. OsGeo Charter Member and currently serving on the Board of Directors. Lead developer at PyCSW, PyGeoapi and PyGeometa. Contact: tomkralidis@gmail.com – GitHub: tomkralidis

  ISO 19115-3 support is valuable to the geospatial community in support of the management, description, and preservation of geospatial data. This support exists in the Geopython community, and would be an asset in the R geometa implementation to achieve similar impact and outcomes.

  Tom Kralidis, Open Source Geospatial Foundation

- Mr François-Xavier Prunayre - Software Engineer, FOSS4G developer, PSC member & committer (GeoNetwork opensource, Talend spatial). GitHub: fxprunayre

  ISO 19115-3 is a major step forward in geographic resources description by bridging various ISO models into a common entity. Data model (aka feature catalogue) can now be described directly

in the metadata. Usage and quality are now 2 specific sections. ISO 19157 is used for quality reports and measures. ISO19115-2 for imagery is now part of the encoding. GeoNetwork is now supporting that standard and some members of the community are adopting it for their daily work eg. Service Public de Wallonie - Metawal, Ifremer - Checkpoints projects, ANZLIC.

User community

- Mr Stephen Roecker, Soil Scientist & GIS Specialist at https://www.nrcs.usda.gov/ – Contact: stephen.roecker@usda.gov:

  The US Geospatial Data Act (2018) is a new US law that requires the federal agencies to publish their metadata on GeoPlatform.gov in ISO 19115. This has focused many folks attention on metadata, like it hasn't been before. Thus, I concur that there is a need for [geometa], and potentially a large user base. This is new territory for most folks, and many would also benefit from user guides.

- Mr Julien Barde - Research Engineer at IRD (French National Research Institute for Sustainable Development) - Contact: julien.barde@ird.fr
- Ms Hélène Raynal - Research Engineer at INRAE (National Research Institute for Agriculture, Food and the Environment) - Contact: helene.raynal@inrae.fr

The Problem

State of development

geometa is a well-established and maintained R package, managed on GitHub, published on CRAN and released in open-source MIT license. It already supports various geographic metadata information international standards such as the ISO 19110 (Feature catalogue), ISO 19115-1 and 19115-2 (geographic metadata for vector and gridded/imagery datasets), ISO 19119 (geographic metadata for service), and ISO 19136 (Geographic Markup Language) providing methods to *read*, *write* and *validate* geographic metadata from R using the ISO/TS 19139 (XML) technical specification.

R native bindings for the ISO 19136 have been enabled to support the export of GML objects from geometry objects modeled with *sf*, also an R implementation of an ISO/OGC standard (ISO 19125 - Simple Features Access). Such native GML support provided by *geometa* is currently unique in the R development community and provides a flexible framework to create objects from the broad GML data model where other libraries (such as GDAL) limit their support to the scope of data feature collections. In addition, the support for GML elements fosters the interoperability between R and standard geospatial web-services that require some GML element bindings as part of their specifications. This is the approach followed with the *ows4R* (OGC Web-Services For R) package, jointly developed together with *geometa*.

Since the first project funded by the R Consortium in 2018, *geometa* has been further consolidated to support various enhancements including features:

- a better support for multilingual metadata, with application in regional and international contexts
- the addition of new classes from the ISO 19136 (GML) and Sensor Web Enablement (SWE) data models, required for enabling standard web coverage service data access (OGC WCS), added in *ows4R* in 2022 (development sponsored by the European Marine Observation and Data Network - EMODnet Biology).

On evolutive maintenance aspects, *geometa* has been evolving to introduce modern R libraries:

- the migration from XML package to the *xml2* R package has been initiated to foster collaboration with active projects, and facilitate the overall maintenance of CRAN packages on the long term.
- the introduction of *waldo* as reference for comparing geometa objects.

Thus, *geometa* provides an essential object-oriented data model in R to allow users managing easily and properly geographic metadata information, in particular by interacting with OGC services and underlying

tools widely adopted by the spatial user community. To fulfill these needs, *geometa* has been exploited in R packages such as *ows4R* (R Interface to OGC Web-Services), *geonapi* (R Interface to GeoNetwork opensource API) and *geoflow* (R engine to orchestrate and run geospatial (meta)data workflows).

Challenge

Despite the above ISO/OGC legacy metadata standards are still widely used by the community, due to the fact that web information systems still extensively use them as main geographic information metadata standards, new standards are set-up by the ISO as part of the continuous standardization process. The ISO/TS 19115-3, produced and approved in 2016 by the ISO TC/221 Technical Committee on Geographic information/Geomatics, was confirmed in 2019. It offers a major evolution of the above geographic information metadata standards by simplifying and consolidating them through an integrated implementation of ISO 19115-1, ISO 19115-2, and concepts from ISO/TS 19139.

Although progressively adopted by several projects from the OpenSource Geospatial Foundation (OSGeo) such as GeoNetwork, PyCSW but also from proprietary software such as ESRI, few are the programmatic tools that allow data managers creating such ISO-19115-3 geographic metadata.

The technical challenge for *geometa* is to re-design its object-oriented model in order to implement the ISO 19115-3 standard while ensuring backward compatibility with previous ISO 19115-1/19115-2/19139 standards.

Stakes are high since *geometa* and the ecosystem of R libraries that is progressively built on top of it, provide a concrete toolset for national and international data managers to take over efficiently the management of geographic information metadata, often constrained by time-consuming and tedious manual editing tasks. At national level, GIS managers are more and more requested by law to manage their geographic information by providing ISO 19115 metadata, such as through the US Geospatial Data Act (2018) or the EC INSPIRE Directive. Backed by these legal frames, and broader objectives such as the UN sustainable development goals, national geospatial data infrastructures are put in place, reinforcing the position of the ISO 19115 as cornerstone of geographic information management. In these infrastructures, the capacity to manage geographic information in R offered by *geometa* in combination with data oriented packages (*sf*, *terra*) is seen as an real asset to implement collective and multidisciplinary works involving data managers, analysts, statisticians and researchers for which R remains the common and most used programming language.

The proposal

Overview

The present project enhances the *geometa* package for handling the new ISO 19115-3 geographic information standard as part of its object-oriented data model developed with R6.

The user community, especially data managers working in research national institutes and international organizations, will take advantage of the features to start adopting the new standard for managing their geographic metadata, progressively promoted in Geographic information management web platforms, especially those from the OpenSource Geospatial Foundation (OSGeo) such as GeoNetwork, PyCSW or GeoNode.

Detail

The object-oriented data model developed with R6 will be enhanced, introducing the following features:

- add mechanism to switch from the ISO 19115-1/19115-2 data model to ISO 19115-3 for handling metadata objects
- adapt existing classes covered by both models to accommodate changes done with ISO 19115-3, and add new classes introduced with ISO 19115-3
- enhance the metadata validator (based on ISO 19115-1/19115-2/19115-3 schemas)

Project plan

Start-up phase

The activities can start immediately as the *geometa* package does already have a GitHub repository. The package has a release cycle in place jointly operated between GitHub, CRAN, and Zenodo for DOI management. The package is released under MIT License. The framework for testing the *geometa* package is already operational, based on GitHub actions.

Technical delivery

The implementation of the ISO 19115-3 standard in *geometa* can be split into several activities / milestones that are listed below together with an effort estimate and an expected delivery date:

- **Milestone 1**: Support for ISO 19115-3 schemas and namespaces; and capacity of *geometa* to switch between standards (ISO 19115-1/2 vs. ISO 19115-3)
  - Effort: 5 days
  - Expected delivery: **2024-07-15**
- **Milestone 2** Enhanced capacity to monitor ISO standards coverage including coverage of ISO 19115-3 schemas in the *geometa* data model
  - Effort: 2 days
  - Expected delivery: **2024-07-31**
- **Milestone 3**: Enhanced capacity to inherit ISO 19115 codelists (through parsing of ISO Codelist catalogues) and application to both ISO 19115-1/2 vs. 19115-3
  - Effort: 3 days
  - Expected delivery: **2024-10-31**
- **Milestone 4**: Evolutive maintenance of ISO 19115 classes already existing in the *geometa* data model for their adaptation to be inteoperable with the ISO 19115-3 standard. This includes the inclusion of additional fields, new methods, and controls over methods field of application (depending on their availability in ISO 19115-1/2 vs. ISO 19115-3)
  - Effort: 10 days
  - Expected delivery: **2024-12-31**
- **Milestone 5**: Support for new classes introduced with ISO 19115-3
  - Effort: 5 days
  - Expected delivery: **2024-12-31**

Jointly with these activities:

- The battery of unit tests will be extended to cover the new standard and ensure backward compatibility with ISO 19115-1/2.
- the R documentation will be progressively updated and checked through quality controls (Github Actions to perform `R CMD Check` in view of releasing the new features on CRAN)
- New features will be directly tested into case studies through collaborations with national and international organizations

Other aspects

Apart from announcements on the package Github page, public advertising of the work will be done through:

- ISC meetings, upon request of the R Consortium ISC for tracking delivery
- Technical oral presentations and/or posters in international conferences (depending on the timeline, features accepted, and available travel funds), such as the useR conference or FOSS4G. Other sources of grants/fundings will be sought to cover eventual conference fees and related travel costs.
- Announcements through the LinkedIn professional network, as public posts on R users groups eg. https://www.linkedin.com/groups/77616/

Requirements

People

Emmanuel Blondel will lead R package enhancements and will involve as required key contributors for specific features of the present proposal, including key developers of the OSGeo community through GitHub.

User community involvement: Outreach will first focus on potential contributors to improve the usability of geometa.

Consulted people for this proposal writing may be contacted to get their feedback during the project implementation, and will be invited to contribute to improve geometa usability, especially in linkage with key OSGeo software projects such as GeoNetwork, PyCSW or GeoNode.

Processes

Source code management

The delivery of the implemented features will comply with R Consortium ISC recommendations. As general process steps, we may mention:

- Issuance of tickets and code commits into the public Github repositories; commits will be referencing ticket numbers for easy tracking of code publication.
- Provision of code documentation including standard R documentation and vignettes.
- Release of package enhancements including Github, CRAN releases, and DOIs (published with Zenodo)

Quality Assurance / Continuous Integration

All implemented features will be subject to a Quality Assurance framework with unit tests (for all R functions) and integration tests (for R functions invoking Web services). These tests will be written according to the *testthat* R package, made publicly available on Github and triggered at each commit through the GitHub platform (Github Actions).

Work progress monitoring

Meetings with R Consortium ISC may occur (upon R Consortium ISC request) to monitor work in progress, and executive summaries may be also be provided to R Consortium ISC to report on the activities

All R codes will be delivered as progressively commited and published in the public Github repository of the *geometa* R package (https://github.com/eblondel/geometa), under a MIT License (set since the package inception).

User community involvement

- In-kind support to users will be done through the Github geometa discussions panel, with a specific topic category set for the newly introduced ISO 19115-3 standard.
- Contributions by anyone to the *geometa* package are open and welcome. For this, contributing guidelines are available at https://github.com/eblondel/geometa/blob/master/CONTRIBUTING.md
- Contributors to the *geometa* project (including the main author/maintainer) are expected to follow the code of conduct available at https://github.com/eblondel/geometa/blob/master/CODE_OF_CONDUCT.md. This code of conduct will be obviously applied for the present project.

Tools & Tech

Tools and Technology required to deliver this project are already available, as follows:

- ISO/TS 19115-3 schemas available at https://schemas.isotc211.org/schemas/19115/

- GitHub repository for *geometa* R package: https://github.com/eblondel/geometa and associated GitHub resources for Continuous Integration tests (GitHub actions)

Funding

The total estimated cost and budget requested here to R Consortium ISC is **13,750 USD** (~12,500 EUR).

Summary

The funding is established based on the time spent by Emmanuel Blondel (Consultant) to carry out the developments as detailed in the *Project plan / Technical delivery* section. Cost is based on an average daily gross rate of 550 USD (500 EUR) and a total of 25 working days (effort estimate).

In-kind contributions are expected to be made in pilot use cases at national and international scale, covering various data domains (earth sciences, oceanography and marine sciences, fisheries, climate change), by collaborating with key contributors as listed in the project *Signatories*.

Success

Definition of done

ISO 19115-3 is integrated into the *geometa* data model, and fully exploitable by the geographic information management user community.

The ISO 19115-3 standard integration in *geometa* is measured through the standards coverage summary available at https://github.com/eblondel/geometa/blob/master/inst/extdata/coverage/geometa_coverage_summary.md as well as the detailed classes/binding inventory at https://github.com/eblondel/geometa/blob/master/inst/extdata/coverage/geometa_coverage_inventory.csv generated automatically by scanning the *geometa* classes & methods and comparing them to the ISO 19115-3 schemas.

Measuring success

Success will be the result of multiple measures, including:

- Package tested code with clear quality assurance indicators made of unit and integration tests
- Package improvements and new features released through CRAN
- Performance and Usability (including available and easy-to-understand documentation)

Future work

Future work and perspectives derived from this project may included:

- the support for ISO 19115-3 automated metadata production through *geoflow* and automated publication into open platforms from the OpenSource Geospatial Foundation (OSGeo) such as GeoNetwork, PyCSW or GeoNode.
- the assessment of the outputs of the RConsortium Object-Oriented Programming Working Group such as the S7 package as possible way to enhance the object-oriented data model implemented in *geometa*.

Key risks

The major technical risk of this project relates to ensuring the backward compatibility of geometa with ISO 19115-1 / 2 standards. This risk will be mitigated by consolidating the battery of unit / integration tests to cover ISO 19115-1, 2 and 3 standards.

In-kind contributions to the present project may be considered from external sources, possibly through handover to some contributors listed in the project *Signatories*, in case some extra effort would be required in relation with the technical delivery effort estimates and expected timeline.